

Edição filológica e digital do Livro do Gado e do Livro de Razão do arquivo do Sobrado do Brejo (Bahia setecentista e oitocentista)

Philological and digital edition of Livro do Gado and Livro de Razão from the Archive of Sobrado do Brejo (Bahia, eighteenth and nineteenth centuries)

Recebido em 31 de março de 2016. | Aprovado em 08 de maio de 2016.

DOI: <http://dx.doi.org/10.17074/lh.v2i1.323>

Mariana Fagundes de Oliveira Lacerda¹
Zenaide de Oliveira Novais Carneiro²

Resumo: Neste trabalho, aborda-se uma documentação de foro privado, o Livro do Gado e o Livro de Razão do Brejo do Campo Seco, no sertão da Bahia, com 57 e 195 folhas, respectivamente, escritos por três gerações, do último quartel do século XVIII ao terceiro quartel do século XIX. São documentos importantes para a história do português brasileiro, de forma particular para a história de penetração e difusão da escrita na Bahia. A edição semidiplomática e edição digital que se pretende realizar desses textos – cujas metodologias estão aqui descritas – atendem a uma das agendas do Projeto Nacional *Para uma História do Português Brasileiro* (PHPB), que vem divulgando *corpora* manuscritos e impressos de períodos pretéritos, para estudo da história do português culto e do português popular do Brasil. Essas edições servirão não apenas a pesquisadores interessados em aspectos linguísticos, da difusão da escrita, da leitura, das transmissões textuais, mas em aspectos históricos, políticos, econômico-sociais, entre outros.

Palavras-chave: livros manuscritos; cultura escrita de foro privado; português brasileiro; séculos XVIII e XIX; edição semidiplomática e digital.

Abstract: In this work, we approach a private forum documentation, the Livro do Gado and Livro de Razão of Campo Seco, in the backlands of Bahia, with 57 and 195 sheets, respectively, written by three generations of the last quarter of the 18th century to the third quarter of the 19th century. These are important documents in the history of Brazilian Portuguese, in a particular way to the history of penetration and diffusion of writing in Bahia. The semidiplomatic editing and digital editing to be achieved in these texts - whose methodologies are described here - attend one of the agendas of the National Project for the Brazilian Portuguese History (PHPB), which has been publishing manuscripts and printed corpora of bygone periods for study the history of cult and popular Portuguese in Brazil. These issues will serve not only to researchers interested in linguistic aspects of writing spread, reading, textual transmissions, but in historical, political, economic and social aspects, among others.

Keywords: manuscript books; private forum culture writing; Brazilian Portuguese; eighteenth and nineteenth centuries; semidiplomatic and digital editing.

¹ Doutora em Linguística pela Universidade Federal da Bahia (UFBA) e vinculada à Universidade Estadual de Feira de Santana (UEFS), na categoria de Professor Adjunto. Atualmente desenvolve pesquisas com apoio da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). marianafag@gmail.com.

² Doutora em Linguística pela Universidade Estadual de Campinas (UNICAMP) e vinculada à Universidade Estadual de Feira de Santana (UEFS), na categoria de Professor Pleno. Atualmente desenvolve pesquisas com apoio da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). zenaide.novais@gmail.com.

Introdução

Numa dimensão histórica, fazendo uma abordagem da história cultural³ no domínio da história da cultura escrita (CASTILLO GOMÉZ, 2003), Lobo e Oliveira (2012, p. 3) afirmam que “reconstruir a história linguística do Brasil passa, também, por reconstruir a história da penetração da língua escrita no Brasil”. A penetração da língua escrita no Brasil consiste na quarta via de investigação apontada por Houaiss (1985)⁴:

4. a penetração da língua escrita no Brasil, das origens aos nossos dias, não numa leitura estética, que se vem tentando algo em vão, nem histórico-externa, nem sociológica, nem demográfica, nem demopsicológica, nem antropológica, nem política, mas essencialmente lingüística – que depois será um componente relevante das “histórias” parciais acima aludidas, cuja conjunção nos possa dar uma história – analítica e sintética – de que já nesta altura tanto necessitamos. (p. 128)

A quarta via em questão coincide com o Campo *a* descrito por Mattos e Silva (1998), que percebe, no âmbito da história do português brasileiro (PB), quatro grandes campos de pesquisa:

a, campo que se moverá na reconstrução de uma história social linguística do Brasil; *b*, campo que se moverá na reconstrução de uma sócio-história linguística ou de uma sociolinguística Histórica; *c*, campo que se moverá na reconstrução diacrônica no interior das estruturas da língua portuguesa em direção ao português brasileiro; *d*, campo que se moverá no âmbito comparativo entre o português europeu e o português brasileiro. (p. 40)

No campo *a*, focalizando a vertente da reconstrução da história da escolarização no Brasil, aborda-se, segundo Mattos e Silva (p. 40), um fator fundamental para compreender a polarização entre normas vernáculas e normas cultas do PB⁵.

Colaborando com o Projeto Nacional *Para uma História do Português Brasileiro* (PHPB), que tem três agendas de trabalho, de acordo com Lobo (2009, p. 306): “constituição de *corpora* diacrônicos de documentos de natureza vária, escritos no Brasil, a partir do século XVI; estudos de mudanças linguísticas depreendidas da análise dos *corpora* constituídos e a reconstrução da história social linguística do Brasil”, propõe-se a edição semidiplomática e digital do Livro do Gado e do Livro de Razão do Arquivo do Sobrado do Brejo, interior da Bahia – dos séculos XVIII e XIX, com 57 e 195 folhas, respectivamente, escritos por três gerações –, focalizando este trabalho, portanto, a primeira agenda, que se refere à formação de bancos de textos.

São apresentados, aqui, os critérios adotados de edição semidiplomática dos referidos documentos, bem como a metodologia de edição digital, usada no âmbito do projeto CE-DOHS – *Corpus Eletrônico de Documentos Históricos do Sertão*⁶, do qual os livros fazem parte. Essas edições servirão não apenas a pesquisadores interessados em aspectos linguísticos, da difusão da escrita, da leitura, das transmissões textuais, mas em aspectos históricos, políticos, econômico-sociais, entre outros.

Trata-se também, na seção 1, mas não de forma sistemática nesta oportunidade, de algumas questões-problemas apresentadas por Petrucci (2003, p. 7-8), para as quais os historiadores da cultura escrita devem voltar-se; são elas:

³ Fala-se aqui em história cultural enquanto uma *História Social da Cultura*, a história cultural propriamente dita, segundo Barros (2004, p. 56), que se distingue de uma *História da Cultura*, limitada a examinar estilisticamente certos objetos culturais pertencentes à “grande Arte” ou à “grande Literatura”.

⁴ A primeira via diz respeito ao levantamento de depoimentos diretos e indiretos sobre todos os processos languageiros havidos a partir dos inícios da colonização (e mesmo antes, para com os indígenas e negros); a segunda via trata do mapeamento confiável da dialectologia brasileira e a terceira via refere-se à dialectologia vertical em tantos quanto possíveis grandes centros urbanos e focos rurais antigos.

⁵ Sobre a polarização e pluralidade do PB, ver Lucchesi (1994).

⁶ O projeto CE-DOHS – *Corpus Eletrônico de Documentos Históricos do Sertão*, financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB)/5566-2010, é desenvolvido no Núcleo de Estudos em Língua Portuguesa (NELP), do Departamento de Letras e Artes (DLA) da Universidade Estadual de Feira de Santana (UEFS), e coordenado por Zenaide de Oliveira Novais Carneiro e Mariana Fagundes de Oliveira Lacerda. Participam do projeto diversos professores, além de mestrands, doutorands e estudantes de iniciação científica. Ver <www.uefs.br/cedohs>.

Qué? En qué consiste el texto escrito, qué hace falta transferir al código gráfico habitual para nosotros, mediante la doble operación de lectura y transcripción. *Cuándo?* Época en que el texto en sí fue escrito en el testimonio que estamos estudiando. *Dónde?* Zona o lugar en que se llevó a cabo la obra de transcripción. *Cómo?* Con qué técnicas, con qué instrumentos, sobre qué materiales, según qué modelos fue escrito ese texto. *Quién lo realiza?* A qué ambiente sociocultural pertenecía el ejecutor y cuál era en su tiempo y ambiente la difusión social de la escritura. *Para qué fue escrito ese texto?*Cuál era la finalidad específica de ese testimonio en particular y, además, cuál podía ser en su época y en su lugar de producción la finalidad ideológica y social de escribir^{7/8}.

Uma análise mais detalhada desses documentos, considerando essas questões, possibilitará aprofundar alguns estudos sobre a formação sócio-histórica do PB no interior da Bahia, resultante de amplo contato linguístico com línguas ameríndias e africanas, entre outras, sobretudo durante o período colonial.

1. Do arquivo do Sobrado do Brejo: o Livro do Gado e o Livro de Razão

O Livro do Gado e o Livro de Razão do Brejo do Campo Seco, no sertão da Bahia, são documentos de foro privado, categoria que tem interessado bastante aos estudiosos ultimamente. Se, antes, de acordo com Castillo Gómez (2003, p. 115), o valor cotidiano da escrita passava despercebido, porque a atenção estava voltada para testemunhos escritos vinculados ao poder, “las orientaciones seguidas en los últimos años han concentrado la atención sobre el valor cotidiano de la escritura”⁹. Na nova história, segundo Burke (1992, p. 11), toda atividade humana interessa, tudo tem uma história; a história da arte, da ciência, da escrita interessam; interessa a escrita de foro privado, não apenas a escrita vinculada à política; a nova história é uma reação deliberada contra o paradigma tradicional, no qual o que tem valor essencialmente é a política.

Esses livros fazem parte do conjunto de documentos que Lobo e Oliveira (2012) situam no Campo 3 de seu projeto de investigação sobre a história da cultura escrita no Brasil: Campo 1: medição de níveis de alfabetismo na história do Brasil; Campo 2: Leitura e escrita aos olhos da Inquisição; *Campo 3: escritas ordinárias e de foro privado na história do Brasil*; Campo 4: prospecção de arquivos e acervos escritos de irmandades negras, de mestiços e de terreiros de candomblé; Campo 5: a escrita do nome próprio de pessoa na história do Brasil; Campo 6: o estudo da língua portuguesa e o fator escolarização em perspectiva histórica; Campo 7: história da cultura escrita das línguas clássicas e de outras línguas no Brasil; Campo 8: fontes iconográficas e manifestações gráficas híbridas (p. 6-12).

O propósito do Campo 3 – escritas ordinárias e de foro privado na história do Brasil –, segundo os autores, é

trazer à luz e analisar as práticas de escrita de cartas particulares, diários íntimos, diários parentais, cadernos escolares, cadernos de confidências, livros de razão e uma infinidade de outras fontes que quase sempre jazeram esquecidas em baús e gavetas, não apenas para os retirar da constante iminência de destruição que sobre eles paira, mas para investi-los do caráter legítimo de objetos de investigação que podem mobilizar, até mesmo em projetos interdisciplinares, historiadores, linguistas, antropólogos etc. (p. 9).¹⁰

O Livro do Gado e o Livro de Razão encontram-se sob a guarda de um familiar, como acontece normalmente com os acervos escritos familiares; o acesso a esses acervos é fundamental para o sucesso do campo 3 referido. Os livros estão hoje em Campinas, São Paulo, sob a guarda de Lycurgo dos Santos Neto, filho de Lycurgo dos Santos Filho, que recebeu os livros de Marieta Betim Pais Leme Canguçu, em Campinas, e Hermengarda Fraga Canguçu, em Bauru, as viúvas, respectivamente, de Artur Gutierrez Canguçu e Inocêncio Canguçu, os netos de Exupério Pinheiro Canguçu, falecido em 1900, o último senhor do Brejo do Campo Seco. Foi com base neles e

⁷ Tradução: “O quê? Em que consiste o texto escrito, que é necessário para transferir o código gráfico habitual para nós, através da operação dupla de leitura e transcrição. Quando? Época em que o texto em si foi escrito no testemunho que estamos estudando. Onde? Zona ou lugar em que se levou a cabo a obra de transcrição. Como? Com que técnicas, com que instrumentos, sobre que materiais, segundo que modelos foi escrito esse texto. Quem o realizou? A que ambiente sociocultural pertencia o executor e qual era em seu tempo e ambiente a difusão social da escrita. Para que foi escrito esse texto? Qual era a finalidade específica desse testemunho em particular e, además, qual podia ser em sua época e em seu lugar de produção a finalidade ideológica e social de escrever.”

⁸ Grifos do autor.

⁹ Tradução: “as orientações seguidas nos últimos anos têm concentrado a atenção sobre o valor cotidiano da escrita.”

¹⁰ Grifos dos autores.

noutros papéis e objetos do arquivo do Sobrado do Brejo aos quais teve acesso que Santos Filho escreveu *Uma comunidade rural do Brasil Antigo: aspectos da vida patriarcal no sertão da Bahia nos séculos XVIII e XIX*, livro publicado em 1956. Em 2012, foi publicada, pela UEFS Editora e pela Fundação Pedro Calmon, a 2ª edição do livro. O projeto CE-DOHS teve a anuência de Lycurgo dos Santos Neto para estudar essa documentação.

Esses dois livros manuscritos do final do período colonial que se conservaram no arquivo do Sobrado do Brejo, da família Pinheiro Canguçu, na fazenda de criação do Brejo do Campo Seco, no povoado de Bom Jesus dos Meiras – que pertenceu à Comarca de Rio de Contas –, hoje denominado Brumado, na região da Serra Geral, no sertão baiano, nos chegam às mãos por meio da leitura e análise de Santos Filho (1956), segundo o qual “Já no século XVIII os criadores do Nordeste baiano possuíam livros para seus próprios assentamentos” (p. 109). Esses livros eram uma prática cultural de escrita naquela época, nas fazendas. Segundo o autor (p. 109), “talvez alguns livros ainda existam, encafuados em arquivos de família nas demais propriedades da região”.

Além dos livros do Brejo do Campo Seco, Algranti (1997, p. 133) refere-se, ainda, como exemplo de escrita doméstica do final do período colonial, ao *Caderno de assentos* do Coronel Francisco Xavier da Costa Aguiar, rico comerciante português, residente na cidade de Santos, em São Paulo, no final do século XVIII. Segundo a autora,

Para o final do período colonial, dispomos de alguns breves exemplos de escritura doméstica, como os *Livros de Razão* e de *Assentos*, uma espécie de memorial que recebia os lançamentos referentes às transações diárias de compra e venda ou troca de produtos, registros de alguns eventos familiares, como nascimento, batizado ou morte dos filhos, de dívidas e prejuízos ocorridos. Uma espécie de protocolo de escritura, no qual o chefe da casa dirige a ele próprio as informações que considera importantes, principalmente para o controle de suas finanças.

Algranti (1997, p. 132-133) ressalta que recuperar o cotidiano no interior dos domicílios coloniais é uma tarefa difícil, porque são extremamente raros registros sistemáticos. Diários íntimos, segundo ela – que poderiam fornecer alguns indícios sobre as atividades desenvolvidas e as relações pessoais havidas ali –, se existiram, não se tornaram públicos, talvez por pudor¹¹.

O Livro de Razão é mais espesso do que o Livro do Gado; este tem 57 folhas, com registros de operações pastoris, de campo, como a marcação de bezerros, de poldros e poldras e de muares, a distribuição de animais pelas várias fazendas etc., e aquele, 195 folhas, com registros de dívidas e empréstimos, de vendas de tecidos, roupas feitas, ferramentas, gêneros alimentícios, armas de fogo, negócios de gado, aquisição de escravos, pagamento de dízimos, gastos com a construção do Sobrado do Brejo, ajuste de empregados, receitas médicas etc. Têm os livros “idêntico aspecto e igual formato, medindo ambos 30 e meio centímetros de altura por 21 de largura” (SANTOS FILHO, 1956, p. 110).

Os livros de razão, que são mais comuns na Europa dos séculos XVII e XVIII do que na América portuguesa¹²,

Em seu aspecto primeiro e mais elementar, são, em geral, um livro de contas; e mesmo quando são mais desenvolvidos, mais elaborados e mais ricos de informações, ainda assim articulam-se e elaboram-se em torno da contabilidade. Escritos do dia-a-dia, na imediata transcrição cotidiana, baseiam-se num esquema simples: o da vida de cada dia em seu ritmo, seus mais prosaicos aspectos materiais, suas atividades mais comuns, registradas numa escritura elementar, em fórmulas que se repetem. (FOISIL, 1991, p. 334)

Os livros não trazem somente registros usuais referentes às transações da fazenda, mas também informações valiosas do cotidiano da fazenda e de seus ocupantes.

O *Livro do Gado* foi inaugurado pelo português Miguel Lourenço de Almeida, primeiro senhor do Brejo do Campo Seco, em 1755, que escreveu, no livro, até 1785; de 1794 a 1822, escreveu, no livro, o brasileiro Antônio Pinheiro Pinto, genro do português Miguel Lourenço de Almeida e segundo senhor do Brejo, homem de negócio e

¹¹ Documentos que sejam mais próximos de uma escrita cotidiana têm especial relevância para a Linguística Histórica, no estudo do português popular brasileiro, sobretudo os que são produzidos por indivíduos que não possuem maior domínio das habilidades de escrita. Barbosa (2006; 2007) ressalta o valor que têm os produtos de inábeis no estudo de sincronias passadas, haja vista o grau de transparência na escrita de dados da oralidade. Para o autor (2007, p. 484) “a inabilidade de reproduzir as soluções mais fonológicas de várias convenções gráficas torna os inábeis em escrita alfabética um grupo mais que desejado pela pesquisa histórica”.

¹² Sobre famílias e vida doméstica na América portuguesa ver Algranti (1997).

familiar do Santo Ofício, proprietário, em 1755, da fazenda do Campo Seco, uma das últimas fazendas do sertão¹³. Segundo Santos Filho (1956, p. 113), “Sua letra é larga e feia e a grafia de muitas palavras denota ausência de certos e elementares conhecimentos linguísticos”. Seria ele, segundo o autor, menos instruído do que Miguel Lourenço de Almeida, seu sogro.

De 1822 a 1832, Inocêncio José Pinheiro ou Inocêncio Pinheiro Canguçu¹⁴ – terceiro senhor do Brejo, filho de Antônio Pinheiro Pinto e neto de Miguel Lourenço –, de temperamento aventureiro, dado ao esbanjamento, diferente do pai e do avô, fez seus registros no *Livro do Gado*, que foi, portanto, escrito por três gerações. Exupério Pinheiro Canguçu, último senhor do Brejo do Campo Seco, homem dado à política, à administração e à vida social, filho de Inocêncio Pinheiro Canguçu, não escreveu nem no Livro do Gado nem no Livro de Razão¹⁵.

Em 1795, Antônio Pinheiro Pinto inaugurou o Livro de Razão do Brejo do Campo Seco. Escreveu Antônio Pinheiro Pinto até 1821 no Livro de Razão. A partir de 1822, após o falecimento de Antônio Pinheiro Pinto, passou a escrever no livro Inocêncio Pinheiro Canguçu. De acordo com Santos Filho (p. 114-115), “Ele pouco escreveu no livro manuscrito” e “A horrível letra e a péssima ortografia de Inocêncio Pinheiro Canguçu revelam que seus estudos foram ainda mais deficientes do que os recebidos pelo seu pai”¹⁶. Data de 1838 seu último assentamento.

Na análise de Santos Filho (1956), tanto Antônio Pinheiro Pinto como Inocêncio Pinheiro Canguçu seriam pouco instruídos; segundo o autor, porque a grafia nos seus registros demonstra ausência de conhecimentos linguísticos elementares. Teriam eles, usando a expressão de Marquilhas (2000), “mãos inábeis”¹⁷ em alguns aspectos? Haveria, na sua escrita, alguns elementos indicadores de que se encontram em fase incipiente de aquisição da escrita? Se forem os escreventes, numa análise qualitativa, segundo Marquilhas, caracterizados como pouco escolarizados, os livros oferecerão possibilidades de uma aproximação do português popular no Brasil do passado, especialmente no interior da Bahia.

Estudando o processo de escolarização no interior da Bahia, na área, como é hoje conhecida, do semiárido, no século XIX, Carneiro e Almeida (2011, p. 546) afirmam que é possível falar de um português semiculto e principalmente de um português popular; o sertão baiano oitocentista tinha baixíssimos níveis de alfabetização,

¹³ Quando Miguel Lourenço de Almeida se instalou nas terras do Campo Seco, situadas na vila de Nossa Senhora do Livramento de Minas do Rio de Contas, capitania da Bahia, a região tinha muita riqueza em ouro e diamantes. A vila, no final do século XVII, era um povoado de bandeirantes paulistas e mineiros. Cresceu e tornou-se freguesia, depois distrito. Hoje, é a cidade do Rio de Contas. A fazenda do Campo Seco passou a chamar-se também Brejo do Campo Seco, nos fins do século XVIII, quando Antônio Pinheiro Pinto, que era seu senhor, construiu o Sobrado do Brejo.

¹⁴ O neto de Miguel Lourenço de Almeida apropriou-se da designação “canguçu”, variedade de uma onça pintada que vagueava na zona da caatinga, onde se localizava o Brejo do Campo Seco.

¹⁵ Exupério Pinheiro Canguçu não escreveu nos livros, mas há muitos escritos seus à disposição de quem puder arrolar os papeis.

¹⁶ Segundo Santos Filho (1946), Antônio Pinheiro Pinto teve dois filhos com Bibiana Maria de Jesus: Inocêncio Pinheiro Canguçu e Zeferina Maria de Santo Antônio. Esta permaneceu, na condição de mulher, sem saber ler e escrever, dedicada apenas às obrigações domésticas, como era comum naquela época. Inocêncio Pinheiro Canguçu, todavia, aprendeu a ler, a escrever, a contar, rudimentos de religião, e também latim. Ele estudou, segundo Santos Filho, “na escola do mestre Tomás José da Costa, um daqueles inúmeros professores particulares que ensinaram as primeiras letras aos meninos e que existiram, no país, desde o início da colonização.” (p. 38) O menino recebeu a mesma instrução um dia ministrada ao pai. Mas, em matéria de ilustração, afirma o autor, o filho perdia para o pai. De acordo com Santos Filho (p. 39), encontra-se, na sua escrita, uma grande quantidade de erros ortográficos, que o pai e o avô não cometeram. Para Santos Filho, Miguel Lourenço de Almeida é mais instruído do que Antônio Pinheiro Pinto, que é mais instruído do que Inocêncio Pinheiro Canguçu.

¹⁷ Segundo Marquilhas (2000), que se fundamenta também na paleografia italiana, os seguintes dados caligráficos, mas não só, caracterizam o que ela chama de “mãos inábeis”, referindo-se a “autores materiais de um texto enquanto falantes estacionados em fase incipiente de aquisição da escrita” (p. 235): “A. *Ausência de cursus*: o desenho autónomo de cada carácter, ou mesmo de traços de cada carácter, decorrente da falta de agilidade dos músculos da mão (...); B. *Uso de módulo grande*: a dificuldade em integrar as letras num módulo pequeno (...); C. *Ausência de regramento ideal*: a incapacidade de respeitar um pautado mental manifesta-se sobretudo na tendência descendente do alinhamento, à medida que se aproxima a margem direita da folha. (...) D. *Traçado inseguro, aparência desenquadrada das letras, rigidez e falta de leveza do conjunto*: a detecção destas características é naturalmente bastante subjectiva. Só o contraste com textos habilmente executados torna possível afirmar que o produto de um principiante as manifesta. (...) E. *Irregularidade da empaginação*: falta proporção entre as margens, as quais podem nem sequer estar definidas, ou relegam a mancha escrita para um extremo da folha. (...) F. *Letras monolíticas*: há um desconhecimento da alografia combinatória dos sinais em contexto inicial, medial ou final. Esta característica relaciona-se com a falta de cursividade e com o desenho autónomo dos caracteres e, por exemplo, impede o <s> de assumir uma forma arredondada em início de palavra gráfica e uma forma alongada nas posições interior e final” (p. 239-240). Santiago (2012) apresenta uma amostra – que faz parte do CE-DOHS – constituída por 91 cartas pessoais, escritas ao longo do século XX, a maioria nas décadas de 50, 60 e 70, por 43 sertanejos oriundos da zona rural dos municípios de Riachão do Jacuípe, Conceição do Coité e Ichu, localizados na região sisaleira do semiárido baiano; trata-se de cartas escritas, como o demonstra a autora, seguindo os critérios de análise de Marquilhas (2000), por mãos inábeis.

segundo as autoras. Nesse período, as escolas eram demasiadamente precárias e seu funcionamento irregular para falar de standardização linguística. No interior da Bahia, conforme Carneiro e Almeida (2011, p. 550), “O processo de standardização linguística anterior ao século XX teria sido bastante localizado e pouco uniforme.”¹⁸.

O último registro feito no Livro de Razão, quase duas décadas depois de Inocêncio Pinheiro Canguçu, é de Miguel Joaquim de Castro Mirante – genro de Canguçu –, “um indivíduo sorumbático e cismador”, de acordo com Santos Filho (p. 115). Na terceira capa, chegou a escrever, depois de revistar o livro: “Vistos e revistos q’ do ao Fogo.” Mas não o fez, felizmente.

Como se tratam de livros escritos por três gerações, há, no que diz respeito a estudos linguísticos dos documentos, possibilidades de fazer uma análise geracional interessante, numa perspectiva gerativa ou sociolinguística, por exemplo. Pode também o Livro do Gado, inaugurado por Miguel Lourenço de Almeida, português, ser explorado nas discussões sobre o português no Brasil. A esse respeito, a tese de Barbosa (1999) traz dados muito significativos, permitindo o estudo de aspectos do português europeu no Brasil, na última década do século XVIII, a partir de dois *corpora*: 93 cartas de comércio escritas no Brasil por portugueses e 14 documentos oficiais da administração pública do Rio de Janeiro.

2. Edição filológica e Edição digital

A aproximação entre o campo filológico e o campo computacional – observada desde a década de 1990 – encontra-se atualmente em plena expansão. O trabalho em ambiente digital no campo da filologia e da linguística histórica tem sido cada vez mais significativo, fazendo surgir, segundo Crane *et al.* (2008), uma nova filologia, a *e-philology*, ou determinando, de acordo com Schreibman *et al.* (2004), o nascimento das Humanidades Digitais.

Considerando as etapas do processo de constituição dos bancos de textos eletrônicos, vencida a etapa de localização e seleção de documentos, eles são transcritos segundo normas filológicas conservadoras e, a partir dessa transcrição, realiza-se, na etapa seguinte, a edição digital, em linguagem XML, finalizando com o preenchimento dos metadados.

No projeto CE-DOHS, os textos-fonte são apresentados em edição semidiplomática¹⁹, segundo as normas de transcrição do PHPB (MATTOS E SILVA, 2001), sendo oferecidas também informações sobre os documentos, sua descrição extrínseca e intrínseca, e, sempre que possível, dados biográficos sobre os autores ou, no caso das cartas, sobre os remetentes e os destinatários, como nome, origem, idade, nível de escolaridade, profissão, estado civil etc.

A codificação dos dados, textuais e extratextuais (ou metadados), é feita com o uso da ferramenta eDictor, o que possibilita a conversão dos textos para diferentes formatos (TXT, XML, HTML) e evita problemas de processamento eletrônico.

As edições filológicas, fidedignas ao texto original, realizadas segundo critérios de transcrição bem definidos, ganham, nos *corpora* digitais, uma versão modernizada, com padronização da grafia, da acentuação e desenvolvimento de abreviaturas, todas as alterações ficando visíveis ao leitor, o que possibilita o controle e mapeamento das intervenções realizadas nos textos, garantindo a recuperabilidade das formas originais. Respeitam-se, entretanto, na edição digital, as mudanças de parágrafo, de linha, as correções do autor, os acidentes do suporte, a orientação da escrita etc. Com isso, oferece-se uma versão eletrônica de textos sem perder o rigor filológico.

¹⁸ Sobre a penetração e difusão da escrita no interior da Bahia ver Carneiro e Almeida (2011a; 2011b), Carneiro (2011), Santiago (2012) e Jesus (2015).

¹⁹ Segundo Borges *et al.* (2012, p. 32), a edição semidiplomática “situa-se entre a interpretativa e a diplomática, sendo sua prática marcada pela ação menos interventiva que a interpretativa e mais interventiva que a diplomática. Comumente, faz-se pela intervenção do editor no sentido de desenvolver as abreviaturas.” Na edição diplomática, faz-se uma pura e simples transcrição do texto antigo dentro do que os modernos caracteres de imprensa e a transcrição paleográfica permitem; na edição interpretativa, “para além da transcrição e da correcção de erros, o editor actualiza a ortografia e elabora notas explicativas de carácter geral”. (DUARTE, 1997, verbete). (MATTOS E SILVA, 2001). São normas de transcrição filológica bem definidas, com o mínimo de intervenção dos editores, procurando conservar a originalidade dos textos, que oferecem importantes indícios à análise linguística; trata-se de uma edição dirigida, sobretudo, a pesquisadores interessados em estudar aspectos linguísticos do documento. Como afirmam Cambraia e Lobo (1995, p. 10), “Editar um texto consiste em escolher-se, do ponto de vista formal, uma maneira particular de o divulgar, a qual será basicamente determinada a partir do público leitor que se queira atingir.”

Procura-se, no âmbito do CE-DOHS, seguir os mesmos critérios de edição digital e de anotação morfossintática que seguem outros projetos de *corpora* eletrônicos, como o projeto *Corpus Histórico do Português Tycho Brahe* (UNICAMP), o projeto *LaborHistórico*, da Universidade Federal do Rio de Janeiro (UFRJ), o projeto *Post Scriptum: arquivo digital de escritura cotidiana em Portugal e Espanha na Época Moderna*, do Centro Linguístico da Universidade de Lisboa (CLUL), o que garante maior praticidade no trabalho e nas consultas e maior integração entre os pesquisadores²⁰.

2.1 - O método fotográfico Lapelinc

O Livro do Gado e o Livro de Razão foram fotografados de acordo com o método fotográfico do *Laboratório de Pesquisa em Linguística de Corpus* (Lapelinc) – da Universidade Estadual do Sudoeste Baiano (UESB) –, que vem sendo desenvolvido desde 2008 (SANTOS; BRITO, 2014), para uso no processo de transposição, para o formato digital, de documentos históricos, a serem disponibilizados em bancos de *corpora* eletrônicos²¹. Segundo Santos e Brito (2014, p. 2),

desde que metodicamente controlada em suas fases de captura, catalogação, edição, armazenamento, e leitura, a Fotografia apresenta-se como forma altamente viável e produtora de digitalização, permitindo à Linguística, ou outra ciência, acessar imagicamente, de modo confiável, o documento não disponível no local da pesquisa.

Trata-se de um importante método de fotografia, cientificamente controlada, na transposição de documentos manuscritos originais em papel para o formato digital, que tem cinco etapas: a) captura de informações da fonte; b) captura fotográfica da imagem do original; c) catalogação das folhas/imagens do documento; d) edição; e) criação de imagens de uso co-indexadas à imagem-original.

Para execução da etapa 2, foi desenvolvida, no Lapelinc, a Mesa Cartesiana, “concebida para ser um equipamento que possibilite, junto com o registro do documento, o registro visual cientificamente controlado de suas características físicas, tais como: tom, cor, tamanho, acervo/arquivo do original físico, tipologia, paginação e sequenciação” (SANTOS; BRITO, 2014, p. 4).

A seguir fotografias do Livro de Razão – o mais espesso dos livros do Arquivo do Sobrado do Brejo –, segundo o método Lapelinc:

²⁰ Por ocasião do *Workshop Construction and use of large annotated corpora*, realizado na UNICAMP, em 2013, pela equipe do projeto *Corpus Histórico do Português Tycho Brahe*, do qual pesquisadores de diversos projetos de *corpora* eletrônicos participaram – entre eles o CE-DOHS –, reafirmou-se a importância de esses projetos seguirem os mesmos padrões de edição digital e de anotação morfossintática, tendo em vista a praticidade do trabalho e a integração dos pesquisadores.

²¹ Os livros foram fotografados por Jorge Viana Santos, da Universidade Estadual do Sudoeste da Bahia (UESB), nos dias 10 e 11 de novembro de 2012, em Campinas, na casa de Lyrurgo dos Santos Neto – que detém a guarda dos documentos –, com sua autorização.

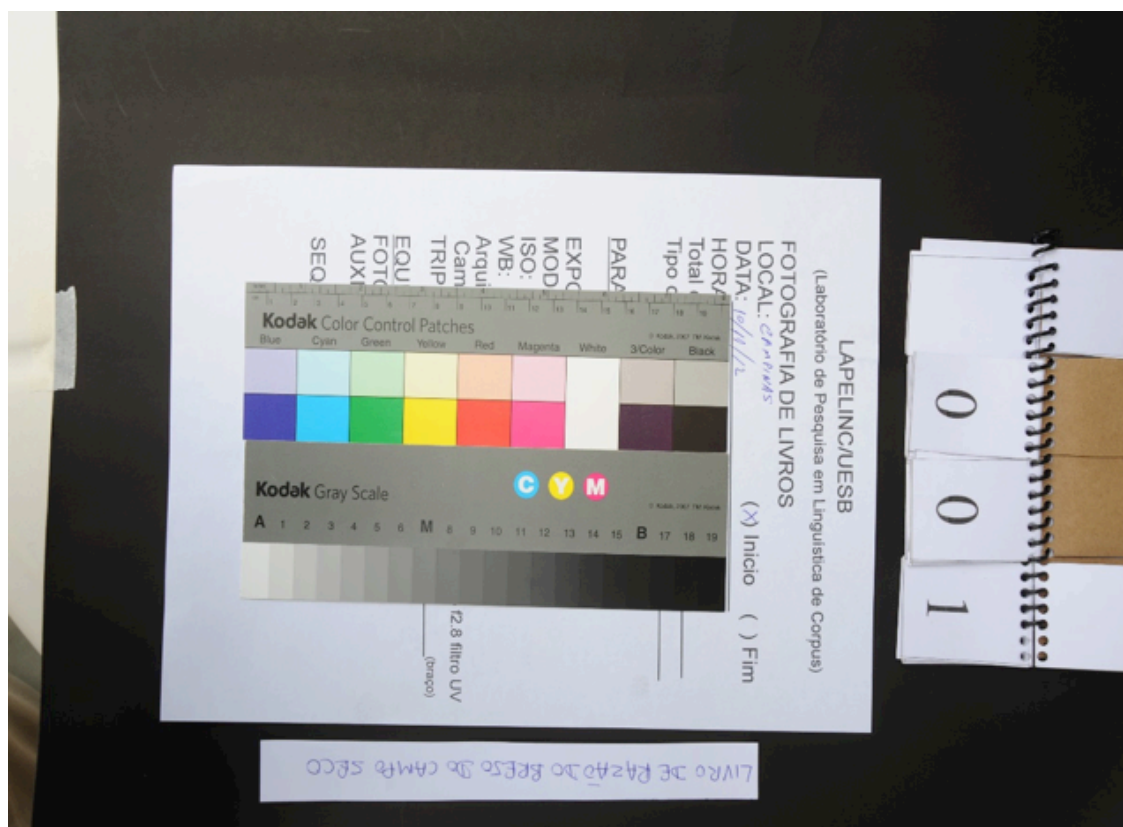


Figura 1. Ficha técnica do Livro de Razão do Brejo do Campo Seco.



Figura 2. Capa do Livro de Razão do Brejo do Campo Seco.

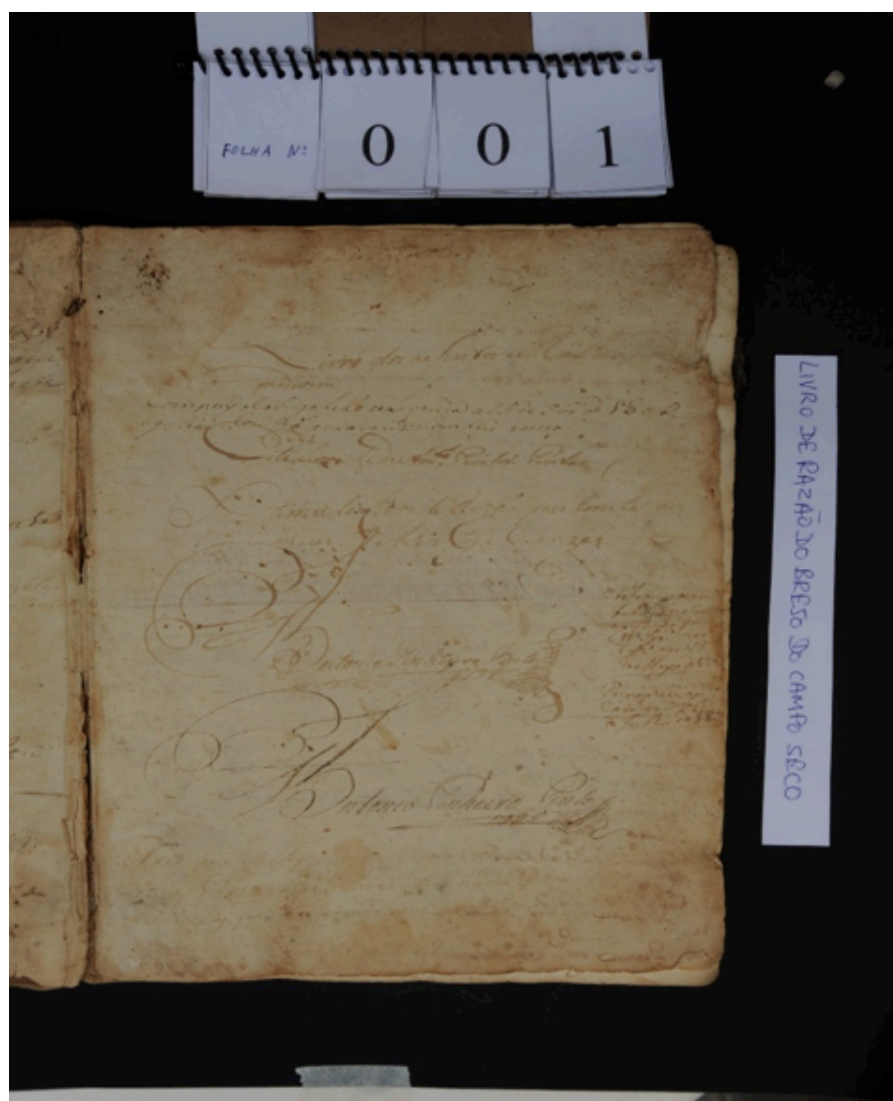


Figura 3. Folha 001 do Livro de Razão do Brejo do Campo Seco.

2.2 - Da edição semidiplomática

No trabalho de edição semidiplomática do Livro do Gado e do Livro de Razão, serão adotadas as normas de transcrição de documentos manuscritos e impressos do PHPB (MATTOS E SILVA, 2001). São normas de transcrição filológica bem definidas, apresentadas durante o II Seminário Para uma história do Português Brasileiro, em Campos do Jordão, em maio de 1998. Há o mínimo de intervenção dos editores, procurando conservar a originalidade dos textos, que oferecem importantes indícios à análise linguística; por exemplo, as abreviaturas, alfabéticas ou não, são desenvolvidas, marcando-se, em itálico, as letras omitidas na abreviatura, obedecendo, entre outros, ao critério de respeitar, sempre que possível, a grafia do manuscrito, ainda que manifeste idiossincrasias ortográficas do escriba; não é estabelecida fronteira de palavras que venham escritas juntas, nem se introduz hífen ou apóstrofo onde não há; a pontuação e a acentuação original são rigorosamente mantidas; é respeitado o emprego de maiúsculas e minúsculas; eventuais erros do escriba ou do copista são remetidos para nota de rodapé, onde se deixa registrada a lição por sua respectiva correção etc. Trata-se de uma edição dirigida, sobretudo, a pesquisadores interessados em estudar aspectos linguísticos do documento. Como afirmam Cambraia e Lobo (1995, p. 10), “Editar um texto consiste em escolher-se, do ponto de vista formal, uma maneira particular de o divulgar, a qual será basicamente determinada a partir do público leitor que se queira atingir.”

2.3 - Da edição digital

A edição digital ou modernizada do Livro do Gado e do Livro de Razão, no âmbito do CE-DOHS, basear-se-á na concepção de hiperedição, que, de acordo com McGann (1997), consiste numa hipermídia capaz de incluir, no mesmo ambiente, edições convencionais (crítica, diplomática, fac-similar, etc.) e diversos tipos de documentos iconográficos, filmográficos, sonoros e textuais, com recursos de zoom, hipertexto e animação. Essa edição será feita com uso do eDictor, desenvolvido por Paixão de Sousa, Kepler e Faria (2007; 2010), um editor de textos especialmente voltado ao trabalho filológico e à análise linguística automática, o qual combina um editor de XML e um etiquetador morfossintático e permite a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em HTML), e de versões com anotação morfossintática (em texto simples e XML).

A padronização ortográfica dos textos originais – que apresentam muitas variações ortográficas – é necessária na edição modernizada, que será anotada linguisticamente, porquanto dessa padronização depende uma maior eficiência de programas de etiquetação automática. Não se costuma intervir, todavia, nos regionalismos, arcaísmos lexicais e neologismos.

Com a ferramenta eDictor, pode-se selecionar a palavra original, que apresenta variação, e editá-la segundo a grafia padrão, ficando disponível a lista de alterações realizadas na edição modernizada. Seleciona-se, manualmente, palavra por palavra que se deseja alterar (substituir, separar, juntar, expandir, etc.)²².

A etiquetagem morfológica e a anotação sintática – feitas com o objetivo principal de possibilitar, de maneira ampla, a recuperação de informações linguísticas dos documentos – são realizadas na edição modernizada dos textos: o programa computacional devolve ao pesquisador, de forma automática, o texto etiquetado, que pode apresentar erros de anotação, os quais devem ser corrigidos pelo linguista, de modo manual.

O CE-DOHS, no processo de anotação linguística dos textos – como o fará no caso do Livro do Gado e do Livro de Razão – segue os mesmos padrões utilizados por outros projetos de *corpora* eletrônicos, a exemplo do projeto *Corpus Histórico do Português Tycho Brahe*, que é o maior *corpus* eletrônico anotado de textos históricos em português.

As etapas básicas que as equipes de pesquisadores de projetos de *corpora* anotados seguem são essas: anotação de edição; etiquetagem morfológica e anotação sintática.

Na primeira etapa, semiautomática, é utilizado o eDictor; trata-se da codificação de informações sobre o texto original, ou sobre decisões editoriais, ou sobre a estrutura do texto.

A próxima etapa, automatizada, é a de etiquetagem morfológica, com uso do programa desenvolvido por Kepler (2007; 2010), um analisador morfossintático automático, com taxa de acerto de 95%, acoplado ao eDictor. Os erros possíveis de etiquetagem, como já dito, devem ser corrigidos manualmente pelo linguista²³.

A anotação sintática, também automatizada, é a terceira e última etapa na constituição de *corpora* anotados; ela diz respeito à identificação e codificação da estrutura sintagmática do texto. É uma tarefa complexa, mais do que a etiquetagem morfológica, e exige um *parser*, ou, na forma portuguesa, um parseador, que realiza a análise sintática, reconhecendo identidades em sequências linearmente dispostas e padrões de agrupamentos hierárquicos.

Considerações Finais

Como disse Mattos e Silva (2002, p. 464), a respeito da reconstrução da história social linguística do Brasil, “O trabalho será muito, de muitos, e não será por pouco tempo.” A edição de documentos e constituição de bancos

²² Os criadores do eDictor, Paixão de Souza, Kepler e Faria (2010), têm trabalhado para sofisticar a ferramenta, tornando-a mais inteligente, a fim de que os processos de edição digital dos textos e sua anotação morfossintática e sintática sejam mais automatizados, facilitando a constituição dos bancos eletrônicos.

²³ O código de etiquetas do eDictor baseia-se no sistema de anotação manual dos *Penn Corpora of Historical English* (KROCH; SANTORINI; DIERTANI, 2010), da Universidade da Pensilvânia, Estados Unidos. Esse sistema, para adequar-se às peculiaridades da gramática do português, sofreu pequenas alterações.

de textos são um passo importante e fundamental nessa empreitada, haja vista que, como afirmou Bacelar do Nascimento (2004, p. 1):

a ampliação de *corpus* favorece essencialmente uma Linguística descritiva, fortemente apoiada pelas novas tecnologias, e permite tomar como ponto de partida da descrição, a análise de quantidade significativa de dados autênticos, à semelhança do que se faz noutros domínios científicos. O uso de *corpora* permite a realização de descrições linguísticas de base empírica e promove, com isso, a discussão de questões teóricas solidamente fundamentadas.”

Hoje, contando com melhores recursos tecnológicos, no universo das Humanidades Digitais, os bancos de textos disponibilizam não somente edições semidiplomáticas, em pdf, mas também edições digitais – a partir do estabelecimento de redes entre projetos que desenvolvem a Linguística de Corpus e a Linguística Computacional –, que servem como recurso eletrônico para estudos linguísticos, entre outros. Como se vê, “Do feliz conagração entre as mais recentes tecnologias e a antiga Filologia, surgiu um novo universo de possibilidades para a preservação, disponibilização e análise de textos antigos, universo em que é possível oferecer ao leitor mais de uma edição do mesmo texto, permitindo que tenha ao seu dispor o texto editado, em diferentes versões, e o seu original.” (GONÇALVES; BANZA, 2013, p. 4).

Os novos recursos que as atuais ferramentas tecnológicas trazem para o campo filológico, com as edições digitais supõem uma verdadeira revolução, tanto em termos de armazenamento textual como em termos de elaboração de *corpora* com diferentes objetivos, como, por exemplo, quantificação de ocorrências, anotação sintática, extração automática das datações lexicais. E ainda, para além dessas possibilidades na análise de textos antigos, como afirmam Gonçalves e Banza (2013, p. 5),

Os acervos ou bibliotecas digitais apresentam-se ao investigador, e bem assim ao público em geral, como verdadeiras bibliotecas globais, possibilitando o acesso a fontes em qualquer lugar do mundo, sem sair do escritório, fontes que podem ser colecionadas num *pen-drive* ou partilhar livremente, com a simples mediação do *notebook* ou de um *tablet*.

O meio digital abre novas possibilidades de trabalho textual, e o projeto CE-DOHS colabora para a edição eletrônica de textos escritos em português²⁴; já disponibiliza diversos acervos, sobretudo de cartas manuscritas, organizando-as por grau de escolaridade e por grau de habilidade com a escrita; são 1084 cartas particulares (1808-2000), num total de 350.850 palavras, escritas por 422 remetentes (nascidos entre 1724 e 1980), extraída a maior parte de Carneiro *et al* (2011). Na segunda fase do projeto, que está em andamento, o número de documentos tem sido ampliado, tanto manuscritos como impressos. Entre os manuscritos, estão os dois livros, aqui apresentados, do arquivo do Sobrado do Brejo.

Atualmente, já se encontram feitas a descrição extrínseca e descrição intrínseca do Livro do Gado e do Livro de Razão; ainda por concluir, a caracterização sócio-histórica da fazenda do Brejo do Campo Seco e a caracterização sociocultural dos escreventes, respondendo, de forma sistemática, às questões-problemas de Petrucci (2003). Em fase de execução, acha-se a edição semidiplomática dos livros, que deverá ser finalizada no primeiro semestre de 2017. A etapa seguinte consiste na edição modernizada dos documentos, usando o eDíctor. Em 2018, as duas edições estarão disponíveis no CE-DOHS.

O CE-DOHS possui um material extenso e rico, que oferece à comunidade científica diferentes possibilidades de pesquisa; e, para a história do PB – especialmente do português no interior da Bahia, através de um contínuo, do mais escolarizado para o menos escolarizado, os *inábeis* –, trata-se de um *corpus* extremamente significativo.

Referências

ALGRANTI, L. M. Famílias e vida doméstica. In: SOUZA, L. de M. e (Org.). *História da vida privada no Brasil: cotidiano e vida privada na América portuguesa*. São Paulo: Companhia das Letras, 1997. p. 83-154.

²⁴ Sobre novas tecnologias que vêm sendo desenvolvidas para a edição eletrônica de textos escritos em português consultar Paixão de Sousa (2013).

- BACELAR DO NASCIMENTO, M. F. *O lugar do corpus na investigação lingüística*. Disponível em: <<http://www.clul.ul.pt/equipa/berlim-2000-nascimento.pdf>>. Acesso em: 20 abr. 2004.
- BARBOSA, A. G. *Para uma história do português colonial: aspectos lingüísticos em cartas do comércio*. Tese de Doutorado. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1999.
- _____. Tratamento dos corpora de sincronias passadas da língua portuguesa no Brasil: recortes grafológicos e lingüísticos. In: LOBO, T. et al. (Org.). *Para a história do português brasileiro*. V VI, t.2. Salvador: EDUFBA, 2006. p. 761-780.
- _____. Normas cultas e normas vernáculas: a encruzilhada histórico-diacrônica nos estudos sobre português brasileiro. In: CASTILHO, A. de et al (Org.). *Descrição, história e aquisição do português brasileiro*. São Paulo: Fapesp, Campinas: Pontes Editores, 2007. p. 483-498.
- BARROS, J. d'A. B. *O campo da história: especialidades e abordagens*. 4 ed. Petrópolis: Vozes, 2004.
- BURKE, P. Abertura: a nova história, seu passado e seu futuro. In: BURKE, P. (Org.). *A Escrita da história: novas perspectivas*. São Paulo: Editora UNESP, 1992. p. 7-37.
- CAMBRAIA, C.; LOBO, T. *Edição diplomático-interpretativa d'A vida do cativo monge confesso*. 1995. Inédito.
- CARNEIRO, Z. O. N. Resultados sobre medição de letramento nos "sertões de dentro" a partir de livros cartoriais (1857-1908). In: *XVI Congresso Internacional da Associação de Lingüística e Filologia da América Latina*. Madrid, 2011.
- _____. (Org.). *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*. Feira de Santana: UEFS, 2011.
- _____.; ALMEIDA, N. L. F. de. A criação de escolas a partir de critérios demográficos na Bahia do século XIX: uma viagem ao interior. In: NEVES, E. F. (Org.). *SERTÕES DA BAHIA: Formação Social, Desenvolvimento Econômico, Evolução Política e Diversidade Cultural*. Salvador: Arcádia, 2011a. p. 603-628.
- _____.; ALMEIDA, N. L. F. Demografia e norma lingüística no semiárido baiano nos séculos XVIII e XIX: uma introdução. In: NEVES, E. F. (Org.). *SERTÕES DA BAHIA: Formação Social, Desenvolvimento Econômico, Evolução Política e Diversidade Cultural*. Salvador: Arcádia, 2011b. p. 599-617.
- CASTILLO GÓMEZ, A. Historia de la cultura escrita: ideas para el debate. *Revista brasileira de história da educação*, n. 5, 2003.
- CRANE, G. et al. *ePhilology: when the boooks talk to their readers*. Blackwell Companion to Digital Literary Studies. Oxford: Blackwell, 2008.
- DUARTE, L. F. *Crítica Textual*. Lisboa: Universidade Nova de Lisboa, 1997.
- FARIA, P.; KEPLER, F. N.; PAIXÃO DE SOUSA, M. C. An Integrated Tool for Annotating Historical Corpora. In: *Fourth Linguistic Annotation Workshop (LAW IV)*, 48th Annual Meeting of the ACL, 2010, Uppsala, Sweden. Proceedings of the Fourth Linguistic Annotation Workshop, 2010. p. 217-221.
- FOISIL, M. A escritura do foro privado. In: ARIËS, P.; CHARTIER, R. (Org.). *História da vida privada: Da Renascença ao Século das Luzes*. V. 3. São Paulo: Companhia das Letras, 1991. p. 331-369.
- GONÇALVES, M. F.; BANZA, A. P. Fontes de metalingüísticas para a história do português clássico. In: GONÇALVES, M. F.; BANZA, A. P. *Patrimônio Textual e Humanidades Digitais: da antiga à nova filologia*. Évora: CIDEHUS, 2013. p. 73-112.
- HOUAISS, A. *O português no Brasil*. Rio de Janeiro: UNIBRADE, 1985.
- KROCH, A.; SANTORINI, B.; DIERTANI, A. *Penn Parsed Corpus of Modern British English*. 2010. Disponível em: [<http://www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/index.html>].
- JESUS, A. S. *Nos caminhos da terra: mensuração dos níveis de alfabetismo no interior da Bahia oitocentista*. Dissertação de Mestrado. Universidade Federal da Bahia, Salvador, 2015.
- LOBO, T. C. F. Arquivos, acervos e a reconstrução histórica do português brasileiro. In: OLIVEIRA, K.; CUNHA E SOUZA, H. F.; SOLEDADE, J. (Org.). *Do português arcaico ao português brasileiro: outras histórias*. Salvador: EDUFBA, 2009. p. 305-327.
- LOBO, T. C. F.; OLIVEIRA, K. *História da cultura escrita no Brasil: um programa de investigação/HISCULTE*, 2012. Inédito.
- LUCCHESI, D. Variação e norma: elementos para uma caracterização sociolingüística do português do Brasil. *Revista Internacional de Língua Portuguesa* 12: 17-28, 1994.
- MARQUILHAS, R. *A faculdade das letras. Leitura e escrita em Portugal no século XVII*. Lisboa, Imprensa Nacional-Casa da Moeda, 2000.
- MATTOS E SILVA, R. V. Idéias para a história do português brasileiro: fragmentos para uma composição posterior. In: CASTILHO, A. T. de. *Para a história do português brasileiro: Primeiras idéias*. V. I. São Paulo: Humanitas, 1998. p. 21-52.

- _____. De fontes sócio-históricas para a história social linguística do Brasil: em busca de indícios. In: MATTOS E SILVA, R. V. (Org.). *Para a história do português brasileiro: primeiros estudos*. V. II, tomos I e II. São Paulo: Humanitas/FFCHL/USP:FAPESP, 2001, v. 2, t. 2, p. 275-302.
- _____. Para a história do português culto e popular brasileiro: sugestões para uma pauta de pesquisa. In: ALKMIM, T. M. *Para a história do português brasileiro: novos estudos*. V. II. São Paulo: Humanitas/FFCHL/USP/ FAPESP, 2002. p. 443-464.
- MCGANN, J. The rationale of hypertext. In: SUTHERLAND, Kathryn. *Electronic text, investigations in method and theory*. Oxford: Clarendon Press, 1997. p. 19-46.
- PAIXÃO DE SOUSA, M. C.; KEPLER, F. N. E-Dictor: Uma ferramenta integrada para a anotação de edição e classe de palavras. *VI Encontro de Lingüística de Corpus*, São Paulo, 2007.
- _____.; KEPLER, F. N.; FARIA, P. P. F. E-Dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: SHEPHERD, T.; SARDINHA, T.B.; PINTO, M. V. (Org.). *Caminhos da linguística de corpus*. Campinas: Mercado de Letras, 2010.
- PAIXÃO DE SOUSA, M. C. A filologia digital em língua portuguesa: alguns caminhos. In: GONÇALVES, M. F.; BANZA, A. P. *Patrimônio Textual e Humanidades Digitais: da antiga à nova filologia*. Évora: CIDEHUS, 2013. p. 113-138.
- PETRUCCI, A. *La ciencia de la escritura: primera lección de paleografía*. Buenos Aires: Fondo de Cultura Económica de Argentina, 2003.
- Projeto CE-DOHS: Corpus eletrônico de documentos históricos do sertão. Disponível em: <www.uefs.br/cedohs>. 2011.
- Projeto *Corpus Histórico do Português Tycho Brahe*. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/>>.
- Projeto *Vozes do sertão em dados: história, povos e formação do português brasileiro*. Coordenação: Zenaide de Oliveira Novais Carneiro. Disponível: <www.uefs.br/nelp>. 2011.
- SANTIAGO, H. S. *Um estudo do português popular brasileiro em cartas pessoais de mãos "cândidas" do sertão baiano*. Dissertação de Mestrado. Universidade Estadual de Feira de Santana, Feira de Santana, 2012.
- SANTOS, J. V.; BRITO, G. S. Fotografia técnica de documentos para formação de corpora digitais eletrônicos: o método Lapelinc. *Revista LETRAS & LETRAS*, v. 30, n. 2, 2014.
- SANTOS FILHO, L. *Uma comunidade rural do Brasil antigo: aspectos da vida patriarcal no Sertão da Bahia, nos séculos XVIII e XIX*. São Paulo: Nacional, 1956.
- SCHREIBMAN, S. et al.. *A Companion to Digital Humanities*. Oxford: Blackwell, 2004.